Mario Castagnino, Luis Lara
Instituto de Física de Rosario
Av. Pellegrini 250, Rosario, Argentina

Olimpia Lombardi
CONICET - Universidad de Buenos Aires.
Puán 470, Buenos Aires, Argentina

A global definition of time-asymmetry is presented. Schulman's two arrows of time model [1] is criticized.

I. INTRODUCTION

The problem of time-asymmetry, also known as the problem of the arrow of time, can be summarized in the following question: *How an evident time-asymmetry is possible if the laws of physics are time-reversal invariant?* In fact, all the laws of physics are invariant under the transformation $t \to -t$[1]. Nevertheless, we have the psychological feeling that past is different than future; moreover, there are clear time-asymmetric phenomena, being the natural tendency from non-equilibrium to equilibrium the most conspicuous example. Astonishing enough, the solution is contained in the above italized question. Since there is a time-asymmetry that cannot be explained by the time-reversal invariant laws (*equations*) of physics, it should be explained by some time-asymmetric *initial conditions.* But, at first sight, initial conditions are arbitrary; therefore, it is impossible to formulate a physical law on initial conditions. However, the initial conditions of any process are the result of another process, in such a way that all processes in a connected universe are coordinated in some way. Therefore, the reason of time-asymmetry is the asymmetry of the universe, namely, a *global* reason. The aim of this letter is to explain this fundamental fact in the clearest possible way, and to discuss some recent criticisms of this global solution [1].

Since Boltzmann's seminal work, many authors had the intuition that time-asymmetry has a global origin. However, traditional discussions usually define time-asymmetry in terms of entropy increasing. In this letter we will reject the traditional entropic approach, following John Earman's [2] "Time Direction Heresy" according to which the arrow of time is an intrinsic, geometrical feature of space-time: this geometrical approach to the problem of the arrow of time has conceptual priority over the entropic approach since the geometrical properties of the universe are more basic than its thermodynamic properties.

II. TIME-ORIENTABILITY

Earman [2] and Grünbaum [3] were the first authors who emphasized the relevance of time-orientability to the problem of the arrow of time. In fact, general relativity considers the universe as a pseudo-Riemannian manifold that may be time-orientable or not. A space-time is *time-orientable* if and only if there exists a *continuous* non-vanishing timelike vector field globally defined. By means of this field, the set of all light semi-cones (lobes) of the manifold can be splitted into two equivalence classes: C_+ and C_-. If space-time were not time-orientable, the distinction between future lobes and past lobes would not be univocally definable on a global level. On the other hand, in a time-orientable space-time, if there were a time-reversal non-invariant law L, defined in a *continuous* way all over the manifold, that would allow us to choose one of the classes as the future one (say C_+) and the other as the past one (say C_-): the law L would be sufficient for defining the arrow of time for the whole universe (namely, a future lobe

---

[1]There are two exceptions:
i) The second law of thermodynamics: the entropy grows. But we use to consider this "law" as an empirical fact that must be demonstrated from more primitive and elementary laws.
ii) Weak interactions. But they are so weak that it is difficult to see how the asymmetry of the universe can be derived of these interactions. Therefore, as it is usual in the literature, we do not address this problem in this paper.

$C_+(x)$ and a past lobe $C_-(x)$ at each point x)[2]. In fact, if one lobe of the class C_+ were considered as the future lobe at a point x and another lobe of the same class were considered as the past lobe at a point y, then joining these two points with a continuous curve (because we only consider connected universes) and propagating the lobe of x towards y (and vice-versa) would be sufficient for finding a point where the law L would be discontinuous, contrary to our supposition.

But, what is this global continuous time-reversal invariant law which allows us to define past and future[3]? This is the essence of Matthews' criticism [5] to the relevance of time-orientability: since there are not continuous and global time-reversal non-invariant laws of nature (but anyway the arrow of time does exists), time-asymmetry is necessarily defined by local laws and, then, it is just a local property; therefore, nothing rules out the possibility that the arrow of time points to different senses in different regions of space-time (see also Reichenbach [6]). What Matthews has forgotten is that an asymmetric *physical fact* can be used to define time-asymmetry instead of a physical law. Of course, it must be an *ubiquitous* physical fact, because it must be used to define the future and the past lobes at all the points of the universe. We will show that the time-asymmetry of the universe is this physical fact by arguing:

1.- That time-symmetric universes belong to a set of *measure zero* on the space of all possible universes.

2.- That the global time-asymmetry of the universe can be used *locally* at each point x to define the future and past lobes, $C_+(x)$ and $C_-(x)$.

We will develop these points in the next two sections.

III. THE CORKSCREW FACTORY THEOREM

In his interesting book, Huw Price [7] emphasizes that time-reversal invariance is not an obstacle to construct a time-asymmetric model of the universe: a time-reversal invariant equation may have time-asymmetric solutions[4]. He illustrates this point with the familiar analogy of a factory which produces equal numbers of left-handed and right-handed corkscrews: the production as a whole is completely unbiased, but each individual corkscrew is asymmetric. Price argument shows the possibility of describing time-asymmetric universes by means of time-reversal invariant laws. But, what is the reason to suppose that time-asymmetric universes have high probability? We will demonstrate that *time-asymmetric solutions of the universe equations have measure zero en the corresponding phase space.*

Let us consider some model of universe equations. All known examples have the following two properties (e.g. see [8], but there are many other examples):

1.- They are time-reversal invariant, namely, invariant under the transformation $t \to -t$.

2.- They are time-translation invariant, namely, invariant under the transformation $t \to t + cont.$[5] (homogeneous time).

To fix the ideas let us consider the simplest example: a FRW universe with radius a and matter represented by a neutral scalar field ϕ. The dynamical variables are $a, \dot{a}, \phi, \dot{\phi}$. They satisfy a Hamiltonian constraint $H(a, \dot{a}, \phi, \dot{\phi}) = 0$ which reduces the dimension of phase space from 4 to 3; then we can consider a phase space of variables $\dot{a}, \phi, \dot{\phi}$. Let us now consider a time-symmetric continuous[6] solution, for $a \geq 0$: there must be a time, that we can take as the time origin $t = 0$ (since the equations are time-translation invariant), with respect to which $a(t) = a(-t)$; therefore $\dot{a}\,(0) = 0$ (e.g., in a big bang-big crunch universe, there must be a maximum; in a universe that begins and ends with a infinite radius there must be a minimum at the bounce of the radius). This means that the radius of the universe is symmetric with respect to $t = 0$. But if we want to obtain the complete time-symmetry of the universe, the field ϕ must be also symmetric with respect to $t = 0$. Since ϕ has not definite sign, we have two possibilities:

---

[2] Of course, a previous requirement is that a cosmic time could be defined in the whole universe, namely, that the manifold would satisfy stable causality condition [4] and that the monotonically increasing function f could be computed from the distance between two hypersurfaces of the corresponding foliation along *any* curve orthogonal to the foliation.

[3] See footnote 1.

[4] Of course, this "loophole" is not helpful when we are dealing with a multiplicity of systems: for each time-asymmetric solution there may be another time-asymmetric solution that is the temporal mirror image of the first one. But when we are studying the whole universe, both solutions are equivalent descriptions of one an the same universe.

[5] We are referring to the equations that rule the behavior of the universe, not to the particular solutions that normally do not have the time translation symmetry.

[6] We will disregard non-continuous solutions since normally information does not pass through discontinuities: we are only considering *connected* universes where information can go from a point to any other time-like connected point.

either $\phi(t) = \phi(-t)$ or $\phi(t) = -\phi(-t)$. Therefore at $t = 0$ we have two possible boundary conditions: $(0, \phi, 0)$ or $(0, 0, \dot{\phi})$ respectively. So the set of $t = 0$ conditions that lead to time-symmetric solutions has dimension 1< 3, and the surface that contains time-symmetric solutions has dimension 2< 3. In the usual Lebesgue measure (or in any measure absolutely continuous with respect to it) both the set of initial conditions that lead to time-symmetric universes and the surface of time-asymmetric solutions have measure zero. q. e. d. .

This theorem can be easily generalized to the case where ϕ has many components, or to the case of many fields with many components where some of these fields may be fluctuations of the metric. Since properties 1 and 2 are also true in classical statistical mechanics, the theorem could be also demonstrated in this case, and also in the quantum case, albeit some quantum gravity problems like time definition [9].

Let us now consider the coarse-grained version of the theorem. Let ε be the size of the grain and, in order to compare measures, let us consider that the phase space $\dot{a}, \phi, \dot{\phi}$ is a cube of volume L^3. In this case, boundary conditions $(0, \phi, 0)$ and $(0, 0, \dot{\phi})$ will be fuzzy and the volume of the set of time-symmetric initial conditions will have measure $2\varepsilon^2 L$. This magnitude can be compared with the size of the phase space, obtaining the ratio $2\varepsilon^2 L/L^3 = 2(\varepsilon/L)^2$. Of course, in the usual case $\varepsilon \ll L$; then, the measure of the set of points corresponding to boundary conditions that lead to time-symmetric universes is extremely smaller than the measure of the phase space. The same argument can be applied to the set of time symmetric solutions with measure εL^2, where $\varepsilon L^2/L^3 = \varepsilon/L << 1$ if $\varepsilon \ll L$. q.e.d.

This completes the first argument announced in Section II; let us now develop the second argument.

IV. THE REICHENBACH-DAVIES DIAGRAM

In his classical book about the arrow of time, Hans Reichenbach [6] defines the future direction of time as the sense of the entropy increasing of the majority of branch systems, that is, systems which become isolated or quasi-isolated from the main system during certain period. Paul Davies [10] appeals to Reichenbach's notion, claiming that branch systems emerge as the result of a chain or hierarchy of branchings which expand out into wider and wider regions of the universe; therefore "the origin of the arrow of time refers back to the cosmological initial conditions".

On the basis of this idea, in previous papers we have introduced the "Reichenbach-Davies diagram" [11], [12], [13][7], where all the local processes which go from non-equilibrium to equilibrium are connected in such a way that the "output" of a process is the "input" of another one: the energy provided by a process relaxing to equilibrium serves to drive another process to non-equilibrium. This "cascade" of processes define a global energy flux which, if traced back, owes its origin to the initial global instability that is the source of all the energy of the universe[8]. On the other hand, the sense of the flux on the time-orientable space-time defines a global *time-orientation*: the incoming flux defines the lobe $C_-(x) \in C_-$ at each point x, the outgoing flux defines the lobe $C_+(x) \in C_+$ at x, and all the lobes of class C_- point towards the initial instability. In this way, the global time-asymmetry of the universe defines the local time-asymmetry in each one of its points.

In summary, the Reichenbach-Davies diagram shows how global time-asymmetry is related with local time-asymmetry, that is, how the different "arrows of time" (cosmological, thermodynamic, quantum, electromagnetic, etc.) are coordinated [12]. The global energy flux is the ubiquitous phenomenon that connects all the processes of the universe. If two sections of the universe are not connected by this flux, then they are completely isolated from each other: each section can be considered as a universe by itself, and within each one of them the global flux is unique. These considerations are particularly relevant for evaluating Schulman's argument.

V. CRITICISM TO SCHULMAN´S ARGUMENT

Schulman [1] exhibits a model in which two weakly coupled systems maintain opposite running thermodynamic arrows of time. From this model, he concludes that regions of opposite running arrows of time at stellar distances from us are possible. This possibility would be a counter-example to our position: Schulman's model would show that a universe consisting in two weakly coupled sub-universes A and B can have two regional arrows of time pointing to opposite senses.

---

[7]At the quantum level it could be considered as the combination of all the scattering processes within the universe. We have called the quantum version of the "Reichenbach-Davies" diagram "Reichenbach-Bohm" diagram [14], [15].

[8]The initial instability is studied in [16] and bibliography therein.

Even though Schulman's argument sounds convincing at first sight, it becomes implausible when analyzed form a cosmological viewpoint. In Schulman's proposal, the low entropy extremities of the sub-universes A and B are opposed, and both sub-universes evolve towards equilibrium in opposed time senses. Let us consider two cases:

1.- The sub-universe A is bigger than the sub-universe B[9] (this situation is not considered by Schulman). If time-orientation is defined by entropy increasing, in this case the time-orientation of the whole universe A∪B will agree with the time-orientation of A, and B will go from equilibrium to non-equilibrium. Nevertheless, the behavior of B is neither strange not unnatural: since there is a flux of energy which, according to the time-orientation adopted, must be considered as a flux from A to B, then we can consider that it is such energy what takes the sub-universe B out of equilibrium. In other words, the decreasing entropy of the *open* sub-universe B has the same explanation as the decreasing entropy in the usual open systems that we find in our everyday life.

2.- The sub-universe A is equal to B (the situation studied by Schulman, where A and B are identical). In this case, the universe A∪B is perfectly time-symmetric. But, as the argument of Section III has shown, time-symmetry has vanishing measure: it requires an overwhelmingly improbable fine-tuning of all the state variables of the universe.

But even in the time-symmetric case, it is not admissible to suppose that the sub-universes A and B have opposite time-orientations. Schulman's cosmological model is a time-orientable universe. In a time-orientable manifold, continuous timelike transport has conceptual priority over any method of defining time-orientation. In other words, Schulman's universe has a light-cone structure such that, if we continuously transport a future pointing vector from the point $x \in$A along some curve to the point $y \in$B, the transported vector will fall into the future lobe $C_+(y)$. This means that A's future cannot be different than B's future: there is an only future for the whole manifold, defined by its light-cone structure.

However, who prefers to insist on the attempt to use entropy increasing for defining time-orientation could appeal to the following strategy: to define the future sense of time as the sense of the entropy increasing, for instance, in the sub-universe A, and then to establish the time-orientation in the sub-universe B by means of continuous timelike transport. But who adopts this strategy is committed to explain why future is the sense of entropy increasing in one region of the universe but not in the other: why the entropy definition works in one region of the universe but not in all of them.

These considerations lead us to our starting point: the problem of the arrow of time should be addressed from a global perspective, taking into account the geometrical properties of space-time.

VI. CONCLUSION.

There is never a last word in physics. But (provisionally) we can conclude that the global definition of the arrow of time has no serious faults and, therefore, it can be used as a solid basis for studying other problems related with the time asymmetry of the universe and its sub-systems.

---

[1] S. Schultman, Phys. Rev. Lett. **83,** 5419, 1999.
[2] J. Earman, Phil. Scie., **41,** 15, 1974.
[3] A. Grünbaum, *Philosophical problems of space and time,* 2nd. ed., Dordrecht, Reidel Pub. Co., 1973.
[4] S. Hawking, J. Ellis, *The large space-time structure,* Cambridge Univ. Press, Cambridge, 1973.
[5] G. Mattews, Phil. Scie., **46,** 82, 1979.
[6] H. Reichenbach, *The direction of time,* University of California Press, Berkeley and Los Angeles, 1956.
[7] H. Price, *Time's arrow and the Archimedes' point,* Oxford University Press, Oxford, 1996.
[8] M. Castagnino, H. Giacomini, L. Lara, Phys. Rev. D, **61**, #107302, 2000, **63**, #044003, 2001.
[9] M. Castagnino, Phys. Rev. D, **29,** 2216, 1989. M. Castagnino, F. D. Mazzitelli, Phys. Rev. D, **42,** 482, 1990. M. Castagnino, F. Lombardo, Phys. Rev. D, **48,** 1722, 1993.

[9]To fix the ideas we can say "externally bigger". The cases "more or less bigger " or "almost symmetric"can be included in the coarse-grained version of the cork-screw theorem, since in these cases there is only a small difference with a symmetric model. Then these solutions have small measure.

[10] P. C. Davies, "Stirring out trouble", in *Physical origin of time asymmetric*, J. Halliwell *et al.* eds., Cambridge University Press, Cambridge, 1994.
[11] M. Castagnino, Phys. Rev. D, **57**, 750, 1998. M. Castagnino, M. Gadella, F. Gaioli, R. Laura, Int. Jour. of Theo. Phys., **38**, 2823, 1999.
[12] M. Castagnino, E. Gunzig, Int. Jour. Theo. Phys., **38**, 47 ,1999.
[13] M. Castagnino, C. Laciana, "The global thermodynamic arrow of time", Class. Quant. Grav., 2002, in press.
[14] M. Castagnino, "The global nature of time asymmetry and the Bohm-Reichenbach diagram", in *Irreversibility and Causality* (Proc. G. 21 Goslar 1996) A. Bohm *et al.* eds., 282, Springer-Verlag, Berlin.
[15] M. Castagnino, A. Ordoñez, J. Math. Phys., **43**, 705, 2002.
[16] R. Aquilano, M. Castagnino, E. Eiroa, Phys. Rev. D, **59,** #087301, 1999.